

SYSTEMS AND METHODS FOR DYNAMIC BANDWIDTH MANAGEMENT ON A PER SUBSCRIBER BASIS IN A COMMUNICATIONS NETWORK

5

CROSS-REFERENCE TO RELATED APPLICATIONS

The present application claims priority from U.S. Provisional Patent Application Serial Number 60/161,182, the contents of which are incorporated by reference.

FIELD OF THE INVENTION

10

The present invention relates generally to bandwidth management in a communications network and, more particularly, to a method and apparatus for providing dynamic bandwidth management on a per subscriber basis in a communications network.

BACKGROUND OF THE INVENTION

15

In order for a host to function properly in a network environment, the host must be appropriately configured. Among other things, this configuration process establishes the protocol and other parameters by which the host transmits and receives data. In one common example, a plurality of hosts are networked to create a local area network (LAN). In the LAN, each host must be appropriately configured in order to exchange data over the network. Since most networks are customized to meet a unique set of requirements, hosts that are part of different networks are generally configured in different manners in order to appropriately communicate with their respective networks.

20

While desktop computers generally remain a part of the same network for a substantial period of time, laptops or other portable computers are specifically designed to be transportable. As such, portable computers are connected to different networks at different times depending upon the location of the computer. In a common example in which the portable computer serves as an employee's desktop computer, the portable computer is configured to communicate with their employer's network, i.e., the enterprise network. When the employee travels, however, the portable computer may be connected to different networks that communicate in different manners. In this regard, the employee may connect the portable computer to the network maintained by an airport or by a hotel in order to access the enterprise network, the internet or some other on-line

25

30

service. Since these other networks are configured somewhat differently, however, the portable computer must also be reconfigured in order to properly communicate with these other networks. Typically, this configuration is performed by the user each time that the portable computer is connected to a different network. As will be apparent, this repeated
5 reconfiguration of the portable computer is not only quite time consuming, but is also prone to errors.

A universal subscriber gateway device has been developed by Nomadix, Incorporated of Santa Monica, California. This universal subscriber gateway is described by United States Patent Applications No. 08/816,174, entitled "Nomadic Router", filed in
10 the name of inventor Short et. al., on March 12, 1997 and No. 09/458,602, entitled "Systems and Methods for Authorizing, Authenticating and Accounting Users Having Transparent Computer Access to a Network Using a Gateway Device", filed in the name of inventor Short et. al., on December 8, 1999. These applications have been assigned to Nomadix, the same assignee of the present invention. The contents of both of these
15 applications are herein incorporated by reference as if fully setforth here within. The gateway device serves as an interface connecting the user/subscriber to a number of networks or other online services. For example, the gateway device can serve as a gateway to the Internet, the enterprise network, or other networks and/or on-line services. In addition to serving as a gateway, the gateway device automatically adapts to the
20 protocols and other parameters of the host, in order that it may communicate with the new network in a manner that is transparent both to the user/subscriber and the new network. Once the gateway device has appropriately adapted data packets transmitted from a host, the host can appropriately communicate via the new network, such as the network at a hotel or at an airport, in order to access other networks, such as the
25 enterprise network, or other online services, such as the internet.

The user/subscriber, and more specifically the remote or laptop user, benefits from being able to access a myriad of communication networks without having to undergo the time-consuming and all-too-often daunting task of reconfiguring their host in accordance with network specific configurations. In this fashion, the gateway device is
30 capable of providing more efficient network access to the user/subscriber. A gateway device is also instrumental in providing the user/subscriber broadband network access

that can be tailored to the user/subscriber's needs. In many instances the remote user/subscriber is concerned with being able to acquire network access and levels of service in the most cost-effective manner. Correspondingly, the gateway device administrator desires the capability to be able to offer the user/subscriber numerous
5 different services and billing rate options. By way of example, the remote user/subscriber in a hotel environment may desire a network subscription for the duration of their hotel stay while the user/subscriber in an airport may desire a network subscription for the duration of their layover or until their scheduled flight departs. Additionally, a user/subscriber may desire a certain level of service based on bandwidth
10 concerns and the need for higher or lower data rate transfers. For example, the user/subscriber who is accessing a network for the purpose of viewing text may desire a lower bandwidth service level that meets their particular needs, however, another user/subscriber who is accessing a network for the purpose of downloading files or media-rich content may desire a higher bandwidth service level capable of transferring
15 data at relatively higher speeds.

In today's fast paced computing and networking environment it is even more advantageous to provide these service and billing options dynamically, allowing the user/subscriber to change, for example, billing rates or bandwidth capacity while a network session is on going. This would allow the user/subscriber to be billed at one rate
20 while downloading the data-intensive file or media-rich content while choosing a more cost-effective billing structure for the less data-intensive activities, such as electronic mail (e-mail). As yet another example, the user/subscriber may be sending and/or receiving data-intensive content which requires a certain minimum transfer rate in order to be effectively communicated at the receiving end, such as voiceover IP or video. In
25 such cases, the user/subscriber may desire some guarantee of delivery of the data-intensive content at a relatively higher cost.

Additionally, the dynamic nature of this process would allow the user/subscriber to change service levels or billing rates without the need to exit the network and initiate a new log-on procedure. In effect, the user/subscriber benefits from having a more
30 efficient and less time-consuming means of altering service levels and billing structure.

However, in current network implementations, bandwidth management is typically performed statically, at the time a new subscriber account is established. The new subscriber typically selects a service level and billing structure defining a particular transmission rate (*i.e.*, bandwidth) when establishing the account, and is not afforded the option of changing their service level or billing structure dynamically. In a common example, the service provider configures the network to provide the new subscriber with the subscriber selected bandwidth service level, which typically involves the service provider configuring hardware devices under the control of the service provider, such as a modem or access controller. The user/subscriber is not allowed to dynamically change the bandwidth, but must contact the service provider and request a change in their configuration to allow a higher or lower transmission rate. In response to such a request, the service provider typically assigns a technician to manually implement the reconfiguration of the hardware device(s) and/or other devices or records. In some instances, the subscriber also may be required to reconfigure their host to accommodate the changes in the bandwidth service level.

Therefore, an unsatisfied need exists in the industry for dynamic management of network access bandwidth, particularly when purchasing network access for a relatively short time, such as in an airport or hotel, or with varying network access speed requirements while at home or in the office.

SUMMARY OF THE INVENTION

The present invention comprises a method and device for dynamic bandwidth management on a per subscriber basis. The user/subscriber can independently set and adjust independently the uplink and downlink bandwidths of their network access service. Thus, the user/subscriber can efficiently manage their network access according to the specific activity on the network.

For example, a user/subscriber can select a relatively high maximum bandwidth for the downlink connection when accessing data-intensive content over the network, or a relatively low bandwidth for the downlink connection when accessing less data-intensive content over the network. The change in bandwidth service level is automatically made substantially instantaneously, and the user/subscriber begins receiving service at the new

bandwidth and, in most instances, at a corresponding new billing rate. Likewise, the user/subscriber can increase or decrease the bandwidth of the uplink connection so the user/subscriber only purchases the amount of bandwidth appropriate for their network activity.

5 In addition, the bandwidth manager provides active management of the delivery of data (also known as and referred to herein as traffic shaping) to increase throughput from the gateway device onto the network. This balances the load on the subsequent link; thereby preventing dropped packets or longer delays.

10 In one embodiment of the invention, a method is defined for dynamic control of data transfer (i.e. bandwidth) by a subscriber in a communications network. The method comprises receiving a data packet at a network device, typically a gateway device and retrieving a subscriber selected bandwidth for the subscriber associated with the data packet. A determination is then made as to whether the transfer rate for data packet transmission should be limited based on the subscriber selected bandwidth. If it is
15 determined that limitation is warranted then the transfer rate for data packet transmission is limited.

20 In another embodiment of the present invention, a method is defined for dynamic control of data transfer (i.e. bandwidth) by a subscriber in a communications network. The method includes receiving a data packet and retrieving a subscriber priority scheme associated with the data packet. A priority determination is then made for transmitting the packet and the packet is queued for delayed transmission.

25 In a further embodiment of the present invention, a device for subscriber bandwidth management is defined as including a bandwidth management module that determines if a received data packet will be delayed from further transmission in order to limit the bandwidth of the subscriber to which the data packet is associated and a queue for queuing the data packet for a delay period if the bandwidth management module determines that a delay period is necessary.

BRIEF DESCRIPTION OF THE DRAWINGS

30 Figure 1 is a block diagram of a communications network incorporating a gateway device in accordance with an embodiment of the present invention.

Figure 2 is a block diagram of a communications network implementing subscriber bandwidth management within a gateway device, in accordance with an embodiment of the present invention.

Figure 3 is a block diagram of a communications network implementing uplink and downlink virtual queues in a gateway device, in accordance with an embodiment of the present invention.

Figures 4A and 4B are related flowchart diagrams of a method for subscriber bandwidth management in accordance with an embodiment of the present invention.

Figure 5 is a flowchart diagram of a method for determining bandwidth for data being transmitted from host to network service, in accordance with an embodiment of the present invention.

Figure 6 is a flowchart diagram of a method for determining bandwidth for data being transmitted from a network service to a host, in accordance with an embodiment of the present invention.

Figure 7 is a flowchart diagram of a method for scheduling data packets for delivery, in accordance with an embodiment of the present invention.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present invention now will be described more fully hereinafter with reference to the accompanying drawings, in which preferred embodiments of the invention are shown. This invention may, however, be embodied in many different forms and should not be construed as limited to the embodiments set forth herein; rather, these embodiments are provided so that this disclosure will be thorough and complete, and will fully convey the scope of the invention to those skilled in the art. Like numbers refer to like elements throughout.

Referring now to Figure 1, a communications network 10 that includes a gateway device 12 is depicted in block diagram form. The network system typically includes a plurality of computers/hosts 14 that access the system in order to gain access to network services or other online services. For example, the hosts can be in communication with ports that are located in different rooms of a hotel or a multi-dwelling residence.

Alternatively, the hosts can be in communication with ports in an airport, an arena, or the

like. The communication network also includes a gateway device that provides an interface between the plurality of hosts and the various networks or other online services. Most commonly, the gateway device is physically located proximate the hosts at a relatively low position in the structure of the overall network system. (i.e. the gateway device will be located within the hotel, multi-unit residence, airport, etc.) However, the gateway device can be located at a higher position in the overall network system such as at a Point of Presence (PoP) or a Network Operating Center (NOC), if so desired.

Although the gateway device can be physically embodied in many different fashions, the gateway device typically includes a controller and a memory device in which commands are stored that define the operational characteristics of the gateway device. Alternatively, the gateway device can be embedded within another network device, such as an access concentrator or a router, or the commands that define the functioning of the gateway device can be stored on a PCMCIA card that can be executed by one or more hosts in order to automatically reconfigure the host(s) to communicate with a different network.

The network system 10 also typically includes an access concentrator 16 positioned between the hosts 14 and the gateway device 12 for multiplexing the signals received from the plurality of communications onto a link to the gateway device. Depending upon the medium by which the hosts are connected to the access concentrator, the access concentrator can be configured in different manners. For example, the access concentrator can be a digital subscriber line access module (DSLAM) for signals transmitted via regular telephone lines, a cable modem termination system (CMTS) for signals transmitted via coaxial cables, a wireless access point (WAP) for signals transmitted via a wireless network, a switch or the like. As also shown in Figure 1, the communication network typically includes one or more routers 18 and/or servers (not shown in Figure 1) in communication with a plurality of network services 20 or other online services 22. While the communication network is depicted to have a single router, the communication network will typically have a plurality of routers, switches, bridges, or the like that are arranged in some hierarchical fashion in order to appropriately route traffic to and from the various networks or other online services. In this regard, the gateway device typically establishes a link with one or more routers. The routers, in turn,

establish links with the servers of other network services or other online service providers, such as Internet service providers, based upon the subscriber's selection.

The gateway device 12 is specifically designed to adapt to the configuration of each of the hosts 14 that log onto the communication network 10 in a manner that is transparent to the subscriber and the network. In the typical communication network that employs dynamic host configuration protocol (DHCP) service, an IP address is assigned to the host that is logging onto the network through communication with the gateway device. The DHCP service can be provided by an external DHCP server 24 or it can be provided by an internal DHCP service located within the gateway device. Once a subscriber opens their web browser or otherwise attempts to access an on-line service, the gateway device will direct the subscriber to enter some form of an identifier such as their ID and password. In an alternate embodiment of the device, it is anticipated that the gateway device will be able to automatically detect this information upon connection of the host to the network or any attempt to log in. The gateway device then determines if the subscriber is entitled to access the communication system, the level of access and/or the type of services to which the subscriber is entitled according to an Authentication, Authorization and Accounting (AAA) procedure that is described by U.S. Patent Application Nos. 08/816,174, 09/458,602 and 09/458,569, previously incorporated by reference. An AAA service, which is a database of subscriber records, may be an AAA server remote to the gateway device or the AAA service may comprise a database incorporated into the physical embodiment housing the gateway device.

Assuming that the subscriber has been authenticated and has authorization, the gateway device typically presents subscribers with a home page or control panel that identifies, among other things, the online services or other communication networks that are accessible via the gateway device. In addition, the home page presented by the gateway device can provide information regarding the current parameters or settings that will govern the access provided to the particular subscriber. As such, the gateway administrator can readily alter the parameters or other settings in order to tailor the service according to their particular application. Typically, changes in the parameters or other settings that will potentially utilize additional resources of the network system will

come at a cost, such that the gateway administrator will charge the subscriber a higher rate for their service.

The home page also permits the subscriber to select the network service 20 or other online services 22 that the subscriber wishes to access. For example, the subscriber
5 can access the enterprise network on which the host is typically resident. Alternatively, the subscriber can access the Internet or other on-line services. Once the subscriber elects to access a network or other online service, the gateway device establishes appropriate links via one or more routers 18 to the desired network or online service.

Thereafter, the subscriber can communicate freely with the desired network 20 or
10 other online service 22. In order to support this communication, the gateway device 12 generally performs a packet translation function that is transparent to the user/subscriber and the network. In this regard, for outbound traffic from the host 12 to the network service or other on-line service, the gateway device changes attributes within the packet coming from the user/subscriber, such as the source address, checksum, and application
15 specific parameters, to meet the criteria of the network service to which the user/subscriber has accessed. In addition, the outgoing packet includes an attribute that will direct all incoming packets from the accessed network service to be routed through the gateway device. In contrast, the inbound traffic from the accessed network service or other online service that is routed through the gateway device, undergoes a translation
20 function at the gateway device so that the packets are properly formatted for the user/subscriber's host. In this manner, the packet translation process that takes place at the gateway device is transparent to the host, which appears to send and receive data directly from the accessed communication network. Additional information regarding the translation function is provided by United States Patent Application No. 08/816,714. By
25 implementing the gateway device as an interface between the user/subscriber and the communication network or other online service, however, the user/subscriber will eliminate the need to re-configure their host 12 upon accessing subsequent networks.

In accordance with the present invention, the gateway device includes a bandwidth manager that enables the user/subscribers at respective hosts to dynamically
30 set and reset the bandwidth of their network access connection. The bandwidth manager can substantially instantaneously adjust the bandwidth available to the user/subscriber

without requiring the reconfiguration or reprogramming of a hardware device(s) such as a network modem or access controller, or the restart of an existing session already in process (i.e. thereby providing session flow control). In addition, the bandwidth manager further enables the user/subscriber to independently set the upstream (uplink to the network) and downstream (downlink from the network) bandwidth. Accordingly, the user/subscriber can establish asymmetric bandwidths for upstream and downstream communications, which may be particularly useful if the user/subscriber's network activity involves a disproportionate amount of data being transferred in one direction (upstream or downstream). Depending on the bandwidth service level chosen by the user/subscriber, the user/subscriber will typically be charged an appropriate service fee. This way, the user/subscriber does not pay for bandwidth that they are not using, and only has to pay for additional bandwidth when requested. In a preferred embodiment, the user/subscriber manages their bandwidth service level via an information and control console as disclosed and described in detail in co-pending United States Patent Application No. 09/541,877, entitled "Information and Control Console for use with a Network Gateway Interface," filed on April 3, 2000, in the name of inventors Short et. al., and assigned to the assignee of the present invention, the disclosure of which is herein incorporated by reference as if set forth fully herein.

In prior art network configurations bandwidth is typically modified statically by having the network administrator establish a baud rate, setting the rate in a modem and the transmitting all data packets in flow at the established baud rate. Additionally, prior art network configurations allow the link level rate to be set through the maximum burst rate or average bit rate. In this fashion, the link level rate in prior art network configurations is typically set statically or manually by a network administrator. The present invention allows for the baud rate to be set at the maximum level and then queuing (i.e. delaying) data packet transmission to simulate a perceived data rate equivalent to a data rate chosen by the subscriber. The gateway device implements the module that allows the subscriber to dynamically adjust the queuing rate of data packets. In this instance, the gateway device can adjust queuing rates on specific packet types (e.g. base queue rates on MAC addresses) and/or specific types of data traffic (e.g. vary queue rates via video traffic versus text). Subscriber profiles that are accessible to the gateway

device are used to determine which data packets are to be queued, the scheduling priority assigned to queued data packets and which packets are to be transmitted without delay.

In accordance with the present invention, the gateway device implements a single queue in each transmission direction (i.e., a single queue for uplink data transmission and a single queue for downlink data transmission). These single queues are responsible for both delaying data packet transmission and prioritizing the transmission of data packets in their respective data delivery direction. In prior art networks data packet queuing is typically implemented in multiple queues; a first queue for classifying the data packets and a second queue for prioritizing the data packets. This type of queuing is processor intensive in that an evaluation of all queues for each packet must be made before a determination of which data packets should be sent based on bandwidth priority. The present invention provides for the use of a ring buffer that is less processor intensive than multiple queue methods. The single queue method of the present invention allows for all data packets in a given slot of the ring buffer to be transmitted simultaneously, thus freeing up the processor until the next tick when the next ring slot of the buffer is processed.

With reference to Figure 2, a block diagram of an embodiment of the gateway device 12 in accordance with the present invention is illustrated. Of particular interest with regard to the present invention is the bandwidth management module 30 (also referred to herein as the bandwidth manager). Typically, the bandwidth manager is implemented by the gateway device or the bandwidth manager can be implemented as a component of the protocol stack of the gateway device. The bandwidth manager processes the data packets (or datagram) passing through the gateway device from the user/subscriber host 14 to the network service 40 and from the network service to the user/subscriber. An intermediary device, such as access concentrator 16 is typically positioned in the network between the hosts and the gateway device. The access concentrator serves to multiplex the signals received from the plurality of hosts onto a link to the gateway device. Depending upon the medium by which the hosts are connected to the access controller, the access controller can be configured in different manners. For example, the access controller can be a digital subscriber line access module (DSLAM) for signals transmitted via regular telephone lines, a cable modem

termination system (CMTS) for signals transmitted via coaxial cables, a wireless access point (WAP) for signals transmitted via a wireless network, a switch or the like.

The bandwidth manager 30 comprises a subscriber selectable bandwidth module 32 and a traffic shaping module 34, both of which are typically implemented in software.

5 The subscriber selectable bandwidth module 32 limits the upstream and downstream bandwidths on each virtual channel through the gateway device 10 to that which the respective user/subscribers selected. The subscriber selectable bandwidth module 32 identifies the associated user/subscriber by the media access control (MAC) address or other suitable subscriber identifier within the data packet. The subscriber selectable
10 bandwidth module 32 communicates with AAA subscriber management interface 50 to retrieve the subscriber profile 52. The subscriber profile will include the user/subscriber selected bandwidth for uplink and downlink data transfer. The subscriber selectable bandwidth module 32 then determines whether or not to reschedule the delivery of that packet at a later point in time in order to prevent the user/subscriber from achieving a
15 bandwidth greater than that which the user/subscriber selected.

The subscriber selectable bandwidth module 32 calculates the appropriate delay, if any, using the size (in bytes) of the current data packet, and the size and time of the previous packet delivered from the subscriber. For example, if the user/subscriber has paid for a downlink bandwidth of 100 kilobits per second (kbps), and the gateway device
20 12 receives a data packet with the size of 1,500 bytes (12,000 bits), it would schedule a delay between packets of 0.12 seconds (12,000 bits in a packet / 100,000 bits per second bandwidth limit).

If it is determined that the packet should be delayed, then the subscriber selectable bandwidth module 32 places the packet in memory in a virtual queue for later delivery.
25 Figure 3 is a block diagram illustrating the use of virtual uplink and downlink queues in a gateway device, in accordance with an embodiment of the present invention. The gateway device 12 will implement a virtual uplink queue 60 and a virtual downlink queue 70. In operation a plurality of subscriber hosts 14, 14n communicate data packets through the communication network 10. The data packets will typically be initially
30 routed through an access concentrator 16 that serves to multiplex the data packets received from the plurality of hosts. The gateway device receives the data packets

coming from the hosts at port 80 and the packets are then communicated to the uplink queue via uplink path 64.

The uplink path is implemented on a per subscriber basis where all traffic from the subscribers is placed into the uplink queue, as required. The subscriber selectable
5 bandwidth module then determines which slot in the queue the data packet is placed in based on how much data from the subscriber is already preoccupying the queue and the degree of utilization of each slot in the queue. If a slot in the queue is full or the subscriber has already placed too much data into a slot, then the data packet will go into a subsequent slot until no more slots are available. At the point at which no more slots are
10 available the queue will drop the packet from further processing. The uplink queue will implement an uplink ring buffer 62 that serves to add delay time to data packets that are placed in the uplink queue. Data packets that leave the ring buffer after the delay period are WAN routed to port 90 for transmission to the network service 40.

The gateway device 12 receives the data packets coming from the network service
15 40 at port 90 and the packets are then communicated to the downlink queue 70 via downlink path 74. The downlink path is implemented on a per subscriber basis where all traffic from the subscribers is placed into the downlink queue, as required. The subscriber selectable bandwidth module then determines which slot in the queue the data packet is placed in based on how much data from the subscriber is already preoccupying
20 the queue and the degree of utilization of each slot in the queue. If a slot in the queue is full or the subscriber has already placed too much data into a slot, then the data packet will go into a subsequent slot until no more slots are available. At the point at which no more slots are available the queue will drop the packet from further processing. The downlink queue will implement a downlink ring buffer 72 that serves to add delay time to
25 data packets that are placed in the downlink queue. Data packets that leave the ring buffer after the delay period are WAN routed to port 80 for transmission to the plurality of hosts 14, 14n.

In a preferred embodiment of the present invention, the virtual queue is implemented by a ring buffer having 120 time slots (set to equal the number of system
30 ticks), each slot being $1/60^{\text{th}}$ of a second. Thus, the ring buffer can delay delivery of a packet up to two seconds. If the delay is to be more than two seconds, then the packet

could be dropped or could replace an existing data packet already in the queue.

Accordingly, the packet is placed in the appropriate timeslot so that the ultimate delivery of a packet will not produce a bandwidth greater than that selected and paid for by the user/subscriber. When the ring buffer pointer is pointing to the timeslot where the packet resides, the packet is transmitted out to the network or out to the user/subscriber, whichever the case may be. Each timeslot of the ring buffer is a link list of packets having a temporary relationship to the other timeslots according to the granularity of a ring buffer. While the ring buffer may have more than 120 timeslots in order to achieve greater accuracy, increasing the granularity produces increased overhead in a system.

10 If the user/subscriber wishes to dynamically change either their downlink or uplink bandwidths, such as via an information and control panel displayed on the host, then the new bandwidth(s) is substantially instantaneously recorded in the user/subscriber's profile. When the next packet having the MAC address of the user/subscriber is processed by the subscriber selectable bandwidth module 32, the new
15 bandwidth is retrieved from the profile and utilized by the subscriber selectable bandwidth module 32 to determine if delivery of the packet should be delayed. Thus, the user/subscriber can dynamically change their bandwidth during an access session without requiring a reconfiguration of a component in the network. Typically, when a subscriber makes a change to the bandwidth limits it will correspond to a different billing rate. This
20 allows the user/subscriber to cost effectively manage their network access service. It is noted that the subscriber selectable bandwidth module 32 typically does not guarantee a minimum bandwidth, but operates to limit the maximum bandwidth available to a user/subscriber according to the upstream and downstream bandwidths selected and paid for by the user/subscriber. In fact, the subscriber selectable bandwidth module 32 may
25 operate to make the user/subscriber's network access slower than it actually would be in the absence of bandwidth management as it queues packets associated with the user/subscriber.

The traffic shaping module 34 of the bandwidth manager 30 functions to limit the number of packets transmitted on the link connecting the gateway device 12 to the
30 network 40. In the embodiment illustrated herein, that link is a wide area network (WAN) link 38. In particular, the WAN link 38 is a communications link of finite

bandwidth capable of only transmitting a certain number of packets over a given period of time. Accordingly, in order to make most efficient the operation of the WAN link 38, it is desirable to limit the packets being transmitted over the WAN link to that which the WAN link can handle without dropping packets. Accordingly, the traffic shaping module 34 uses a virtual queue in both the uplink and downlink directions that operates essentially the same as the virtual queues of the subscriber selectable bandwidth module 32, with the result of preventing the WAN link 38 from becoming overloading. By queuing the packets for a short period of time before transmitting them when the WAN link reaches its bandwidth capacity, the overall efficiency of the data transfers over the WAN link improves because the packets are not dropped and, therefore, do not need to be retransmitted. The traffic shaping module 34 monitors the traffic over the WAN link 38, and when it appears to be close (slightly above or slightly below) to the bandwidth capacity of the link, then the traffic shaping module 34 begins queuing the packets for short periods of time with the expectation that the packet volume will at some point drop so that the packet will not need to be queued for the time being. Thereby, the likelihood, or at least the number, of packets being dropped is minimized.

Another aspect of traffic management performed by the traffic shaping module 34 is the prioritization of services provided to the users/subscribers. For example, in a preferred embodiment, each user/subscriber pays for a class of service which has associated with it a maximum bandwidth. Currently, those users/subscribers which request a relatively large bandwidth compared to another user/subscriber will have its packets given priority over the packets of the other user/subscriber at the point of being transmitted over the WAN link 38. This is achieved by queuing the packets of the other user/subscribers and not those with priority. This prioritization of the packets of one user/subscriber over that of another user/subscriber can be extended to numerous other implementations. For example, a user/subscriber may purchase a class of service that places a higher prioritization upon a particular type of content, for instance, video data, audio data, etc. In such cases, the traffic shaping module 34 looks into each packet to determine the user/subscriber associated with that packet based upon the MAC address,

It is also possible, and within the inventive concepts herein disclosed to have the traffic shaping module 34 look at the packet header and/or IP header to determine the

content of the packet and prioritizes accordingly. Specifically, the packets including a protocol utilized to transmit audio content, for example, voiceover IP, may be given priority by the traffic shaping module 34 if the associated user/subscriber paid for such a class of service. Other protocols/parameters on which priority can be based include

5 SMTP, EDP, TCP, a specific IP address, etc. As such the traffic shaping module can be modified to base priority on packet attributes, protocol type and/or destination port addresses.

Yet another example of the prioritization that can be performed by the traffic shaping module 34 in accordance with the present invention is the allocation of

10 bandwidth based on a subscriber's quality of service. This type of allocation would typically guarantee a minimum level of bandwidth for the subscriber. For example, by monitoring the throughput of the gateway device 12, the traffic shaping module 34 can prioritize the packets of a specific user/subscriber who is paying for a specified percentage of the available bandwidth so that the packets of that user are transmitted at

15 the appropriate rate so as to utilize the percentage of available bandwidth. These and other prioritization and scheduling schemes for traffic shaping are often referred to in the industry as weighted fare queuing, deficit round robin, random early detection or the like.

With reference to Figure 4A, a flow diagram of a method in accordance with an embodiment of the present invention for bandwidth management is illustrated. At block

20 200, a subscriber establishes an authorization account for accessing a communication network through a gateway device in accordance with the present invention. The authorization account typically includes a predetermined bandwidth value, preferably for the uplink and downlink connections, and an authorization code, such as a user name and password. At block 210, the subscriber logs into the gateway device and is authorized

25 based on the subscriber's authorization code. This establishes network access for the subscriber through the gateway device at the predetermined bandwidth found in the authorization account. Next, at block 220, bandwidth management is performed on data packets passing through the gateway device to limit each subscriber's bandwidth to that which they have dynamically selected, and to perform traffic shaping functions, as

30 described in greater detail with the reference to Figure 4B. Managing the bandwidth entails receiving a request from the subscriber to dynamically adjust the predetermined

bandwidth value to a second bandwidth value and adjusting the limit on the information transfer rate to the second bandwidth value. The adjustment may be for the uplink bandwidth, the downlink bandwidth or both the uplink and downlink bandwidths.

For purposes of clarity, it is noted that that the predetermined bandwidth
5 requested by a subscriber, when establishing a network access service, typically a maximum bandwidth, represents merely a predetermined transfer rate, typically a maximum transfer rate, not necessary the actual transfer rate obtained by the user/subscriber. In fact, the user/subscriber will often find that their transfer rate is well below that of their selected rate due to network delays and congestion. Yet at other
10 times, when the network resources are underused, the user/subscriber may obtain a transmission rate close to or at their selected transfer rate.

In Figure 4B, the operation of an embodiment of the bandwidth manager on the data packets being delivered upstream to the network is provided. In particular, at block
15 300, a new data packet is received for processing at a gateway device or similar network interface. At block 310, the data packet is processed by extracting the MAC address from the data packet and retrieving the authorization file associated therewith, preferably from a hash table embodied with a AAA service. Based upon (a) the predetermined bandwidth chosen by the subscriber as determined from the authorization file; (b) the size of the current data packet; and/or (c) the size and time of the previous packet sent by the
20 subscriber and processed at the bandwidth manager, it is determined if the packet needs to be queued for a period of time to ensure that the subscriber does not receive a bandwidth greater than that which the subscriber selected, as determined at decision block 320. If the packet should be delayed, then at block 330, the appropriate delay is calculated and the packet is placed in the appropriate timeslot of a ring buffer. When the
25 pointer of the ring buffer addresses the timeslot in which the packet resides, then the packet is further processed by the traffic shaping module of the bandwidth manager. In particular, at block 340, it is determined if the packet needs to be queued for traffic shaping purposes. If the packet needs to be queued, then it is determined at block 350 how long the packet should be delayed, and then the packet is placed in the appropriate
30 timeslot of the ring buffer. When the pointer of the ring buffer addresses the timeslot where the packet resides, then the packet is transmitted over the network at block 360.

Figure 5 is a detailed flow diagram of the operation of the bandwidth manager in preparing data packets for transmission to an upstream network, in accordance with an embodiment of the present invention. At block 400 the time at which the data packet arrives at the gateway device is determined. The time of arrival is necessary if delay periods are to be determined and implemented for a given data packet. At block 410 the gateway device determines whether the data packet is bandwidth limited. If a determination is made that the data packet is not bandwidth limited then, at block 420, the transfer rate is set to zero so that no delay period is assigned to that data packet. If a determination is made that the data packet is bandwidth limited then, at block 430, the MAC address of the data packet is determined so that the data packet can be associated with the host from which it was sent. In this fashion, the subscriber, or more specifically the subscriber's host, can be identified.

At block 440, the gateway device accesses a database, typically an AAA service, to obtain a subscriber profile associated with the MAC address. An initial determination is made, at block 450, as to whether a profile exists in the database associated with the MAC address. If no profile is found in the database then an error message is returned to the subscriber and the default mechanism sets the transfer rate to the default rate, in accordance with block 420. If a profile is found in the database then, at block 460 the subscriber's uplink transfer rate is obtained from the profile. The gateway device then makes a determination, at block 470, as to whether there is an entry in a short term memory module that tracks the subscriber's current uplink transfer rate. In the instance where no entry currently exists, typically this entry is the first or the operation has been timed out, then, at block 472, an entry in the memory module, typically a hash table, is created. Once a determination is made that an entry already exists in the memory module or an entry is created then, at block 480, the transfer rate is set and, at block 490 the data packet is scheduled for transmission in accordance with the transfer rate that has been set.

Figure 6 is a detailed flow diagram of the operation of the bandwidth manager in preparing data packets for transmission from the network to the downstream host, in accordance with an embodiment of the present invention. At block 500 the time at which the data packet arrives at the gateway device is determined. The time of arrival is

necessary if delay periods are to be determined and implemented for a given data packet. At block 510, the destination MAC address of the data packet is determined so that the data packet can be associated with the host from which it will be sent to. In this fashion, the receiving subscriber, or more specifically the subscriber's host, can be identified. At block 520 the gateway device determines whether the data packet is a broadcast (i.e. audio, video or the like) data packet. If a determination is made that the data packet is a broadcast data packet then, at block 530, the transfer rate is set to zero so that no delay period is assigned to that data packet.

If a determination is made that the data packet is not a broadcast data packet then, at block 540, the gateway device accesses a database, typically an AAA service, to obtain a subscriber profile associated with the destination MAC address. An initial determination is made, at block 550, as to whether a profile exists in the database associated with the destination MAC address. If no profile is found in the database then an error message is returned to the subscriber and the default mechanism sets the transfer rate to the default rate, in accordance with block 530. If a profile is found in the database then, at block 560 the subscriber's downlink transfer rate is obtained from the profile. The gateway device then makes a determination, at block 570, as to whether there is an entry in a short term memory module that tracks the subscriber's current downlink transfer rate. In the instance where no entry currently exists, typically this entry is the first or the operation has been timed out, then, at block 572, an entry in the memory module, typically a hash table, is created. Once a determination is made that an entry already exists in the memory module or an entry is created then, at block 580, the downlink transfer rate is set and, at block 590 the data packet is scheduled for transmission in accordance with the transfer rate that has been set.

Figure 7 is a detailed flow diagram of the process by which data packets are queued for transmission as a means for providing bandwidth management to a subscriber in a communications network, in accordance with an embodiment of the present invention. At block 600 a determination is made whether the ring usage exceeds the maximum ring usage capacity. If the maximum ring usage capacity is exceeded (i.e. the ring is overflowed) then, at block 610, the packet is dropped and no further processing occurs. If the maximum ring usage capacity is not exceeded the, at block 620, a

determination is made as to whether the packet has unlimited bandwidth. For example, media packets will typically be designated as having unlimited bandwidth. If the packet is determined to be designated as having unlimited bandwidth then, at block 630, the data packet is queued for immediate transmission by setting the slot to the current slot being transmitted. If the determination is made that the data packet is not designated for unlimited bandwidth then, at block 640, a calculation is made to quantify the difference between the bandwidth available to the subscriber and the bandwidth currently being used (i.e. the remaining available bandwidth for the subscriber).

At block 650, it is determined if there is room in the ring buffer for the data packet. If there is no room in the ring buffer then the packet is dropped at block 610. If there is room then, at block 660, the bandwidth management module initiates looking at the first available slot in the ring buffer with available bandwidth. At block 670, a determination is made as to whether allocating the data packet to ring buffer slot that is currently being looked at would overflow the bandwidth limit of that particular slot. If it does not overflow the slot then, at block 680, the data packet is inserted into the slot. If it would overflow the slot then, at block a determination is made as to whether more slots are available. If no more slots are available then, at block 610, the data packet is dropped (i.e. no further processing). If further slots are available then, at block 700, the pointer advances to the next available slot in the ring buffer. An overflow assessment is made at block 670 and the iterative process continues until the data packet is placed in a slot with sufficient available capacity or until the data packet is dropped.

The subscriber bandwidth management process and device of the present invention allows users/subscribers in a communications network to dynamically alter bandwidth limits independently in both the uplink and downlink data transmission paths. This is accomplished by providing for a single queue in the uplink transmission path and a single queue in the downlink transmission path. Thus, the user/subscriber can efficiently manage their network access according to the specific activity on the network. The network manager benefits from being able structure bandwidth allocation on a per subscriber basis so that overall data transmission is made more efficient. In addition, the bandwidth manager provides active management of the delivery of data (also known as

and referred to herein as traffic shaping) to increase throughput from the gateway device onto the network.

Many modifications and other embodiments of the invention will come to mind to one skilled in the art to which this invention pertains having the benefit of the teachings presented in the foregoing descriptions and the associated drawings. Therefore, it is to be understood that the invention is not to be limited to the specific embodiments disclosed and that modifications and other embodiments are intended to be included within the scope of the appended claims. Although specific terms are employed herein, they are used in a generic and descriptive sense only and not for purposes of limitation.